

Los datos en el contexto actual: Análisis de aplicaciones

V Festival de los Libros
III Jornadas del conocimiento
“El mundo de los datos”

SARA DIAGO GONZALVO
2 al 4 de noviembre de 2017
Biblioteca EPM
Medellín, Antioquia - Colombia

- **¿Qué información se considera valiosa en la búsqueda de soluciones frente a las problemáticas emergentes?**
- **¿Cuáles son las herramientas más útiles para ello?**
- **¿Cómo analizar y gestionar la enorme cantidad de datos digitales que estamos generando?**
- **¿Cómo y para qué implementar Open Data en las bibliotecas?**
- **¿Cuál es la información que necesitamos para impulsar el desarrollo global?**

¿Qué información se considera valiosa en
la búsqueda de soluciones frente a las
problemáticas emergentes?

Agentes

Ciudadanía



Economía – Sector infomediario



Estado

Ciudadanía

- Información económica, presupuestaria y estadística
- Información institucional, organizativa y de planificación
- Información de relevancia jurídica

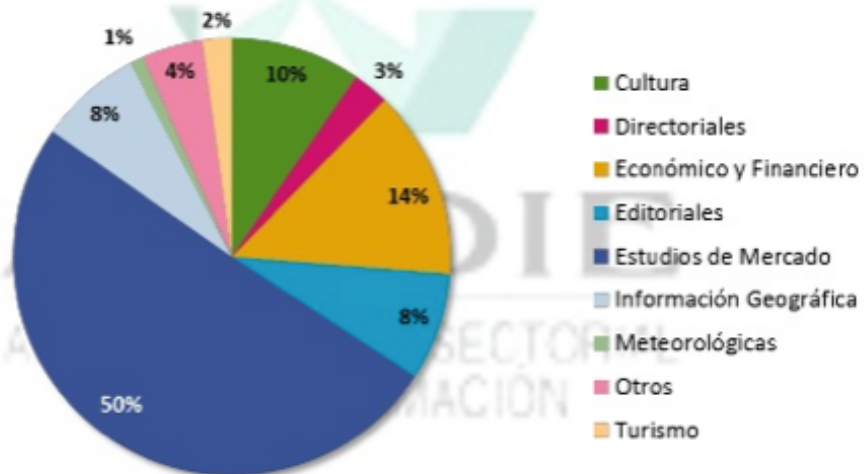


Transparencia en la gestión pública

**Demandante de información pública.
Comprensión
Monitorización**

Economía – Sector infomediario

Distribución de sociedades por sector de actividad

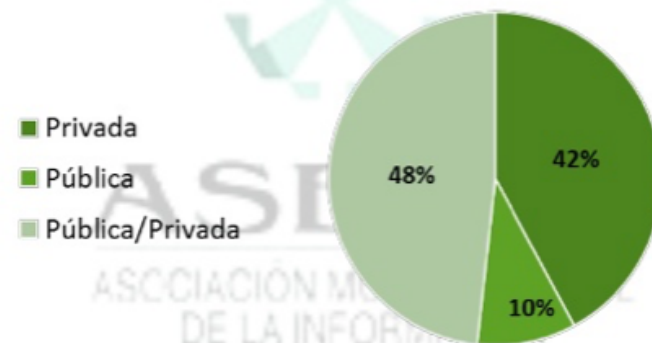


Fuente: Informe ASEDIE* 2016



**Dataset normalizados
Reutilización**

Distribución de sociedades por tipo de
información utilizada

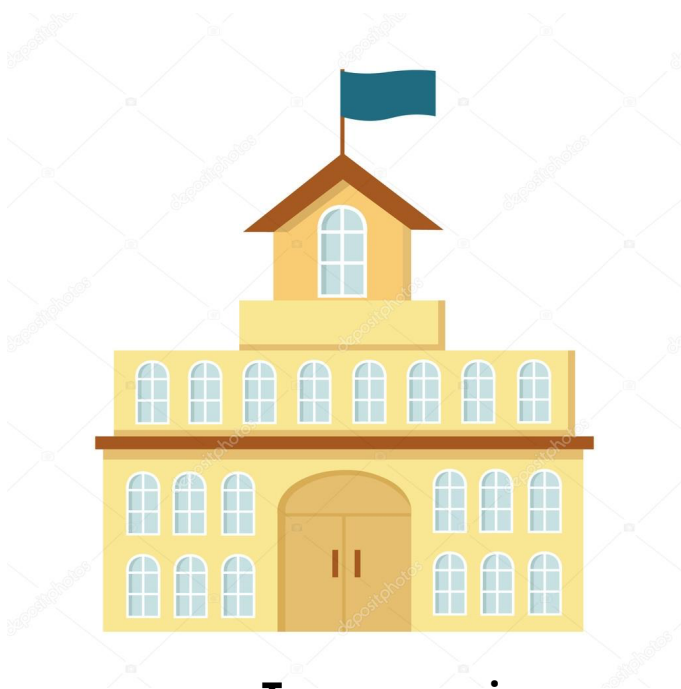


Fuente: Informe ASEDIE* 2016

*ASEDIE: Asociación multisectorial de la información

Estado - Gobierno

- Regulación normativa
- Open Data
- Apertura
- Reutilización
- Prevención del fraude



Transparencia
Eficiencia
Alertas

Datos de la gestión pública

Proyecto 1

Análisis de los Altos Cargos de la Generalitat Valenciana



Juan S. Bejarano Ballén

Master MUGI

***Ingeniero informático con especialización en
Negocios y Servicios en Telecomunicaciones en la
Universidad de SanBuenaventura***

Nacionalidad: Colombia



@juabebal

Proyecto 2

La contratación pública al detalle



Sara Diago Gonzalvo



Master MUGI

Grado en Administración y Dirección de Empresas y

Diplomada en Ciencias Empresariales por la

Universidad de Valencia

Nacionalidad: España

 ***@sardiagon***

**¿Cuáles son las herramientas más útiles
para ello?**

Herramientas para el desarrollo

Tesseract OCR

Obtiene caracteres como imagen para quitar protección.

Realiza tareas de extracción de información a gran escala.

spaCy



elasticsearch

Servidor de búsqueda, permite operar sobre múltiples índices a la vez

Grafo (Visualiza, analiza y realiza búsquedas)



neo4j

Análisis de los Altos Cargos de la Generalitat Valenciana*

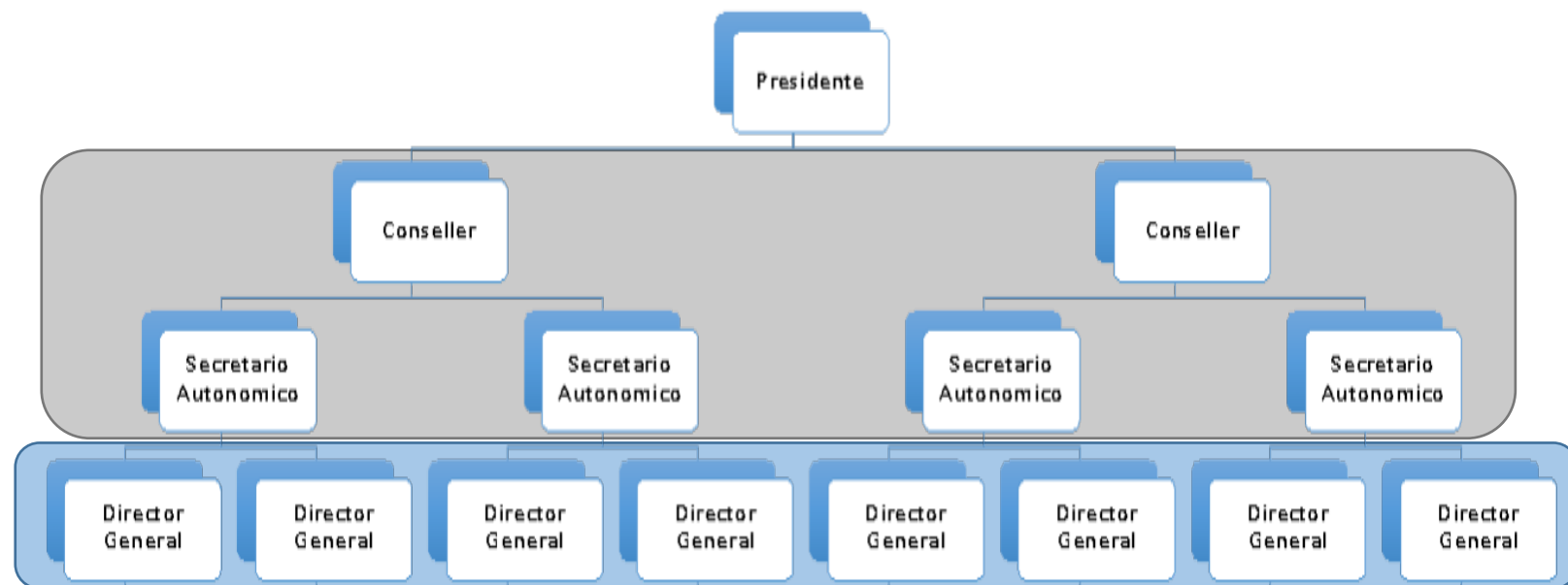
- Juan Sebastián Bejarano Ballén

*Generalitat Valenciana equivalente a Gobernación del Departamento en Colombia

Paso 1

Órganos Superiores

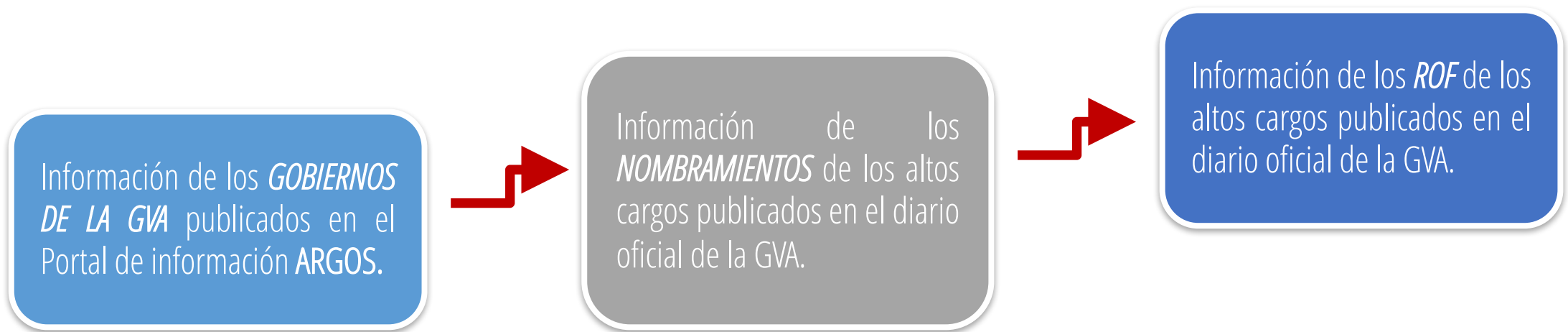
Nivel Directivo



Nivel Administrativo

Fuente: Elaboración propia

Paso 2



Fuente: elaboración propia

Información extraída de nombramientos	1292 ficheros
Fecha de ficheros de altos cargos extraídos.	02/08/1979 – 19/05/2017
Fecha de Documentos usados	21/06/2003 – 19/05/2017
Información extraída de los ROF	407 ficheros
Total de Consellerías Analizadas	39
Número de Consellers	110
Número de Secretarios Autonómicos	86
Número de Subsecretarios	40
Número de Directores Generales	231

Fuente: Elaboración propia

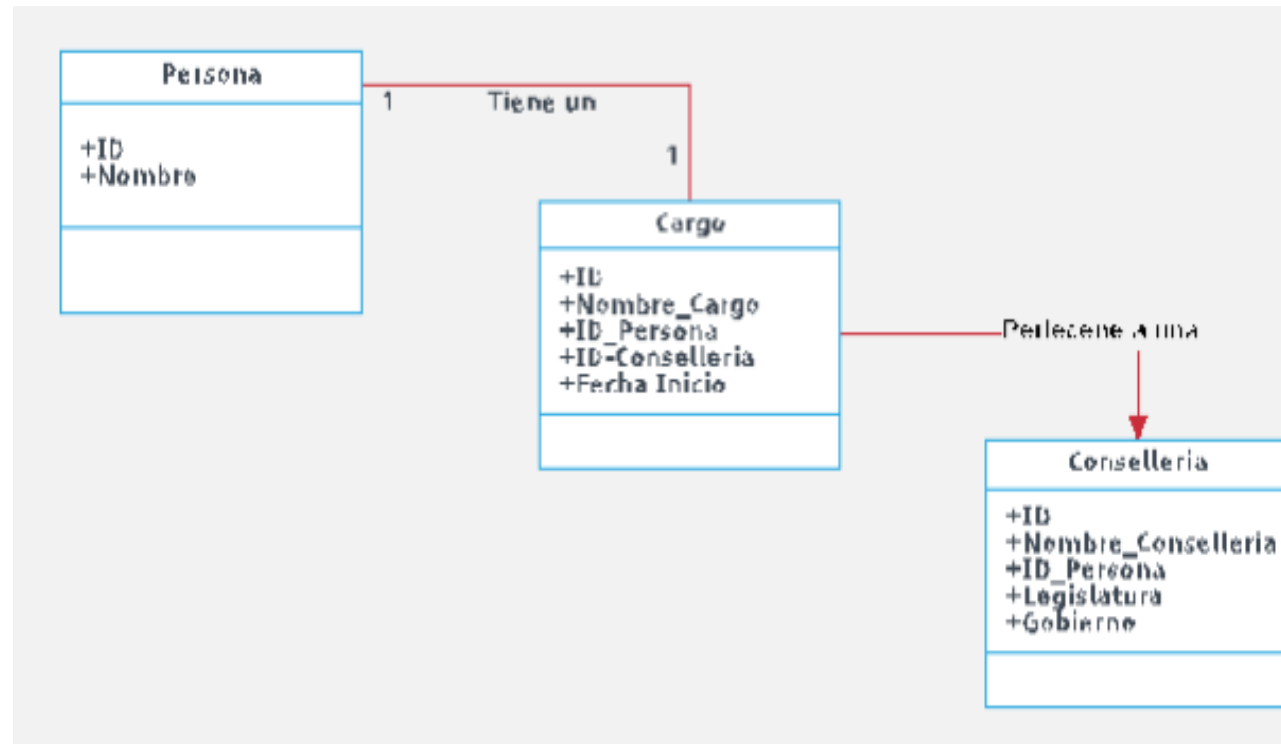
Minería de datos

La minería de datos (Data Mining) es el proceso llevado a cabo para detectar información en grandes conjuntos de datos, hechos a medida que describen características de objetos, eventos o personas, utilizando análisis matemático (algoritmos)



Fuente: adaptada de Introducción a la minería de datos

Paso 3: Diagrama de relación



Fuente: Elaboración propia

Paso 4: Generar ficheros

FICHERO 1

INFORMACIÓN DE LOS
ALTOS CARGOS



FICHERO 2

CONSELLERIAS /
ROF

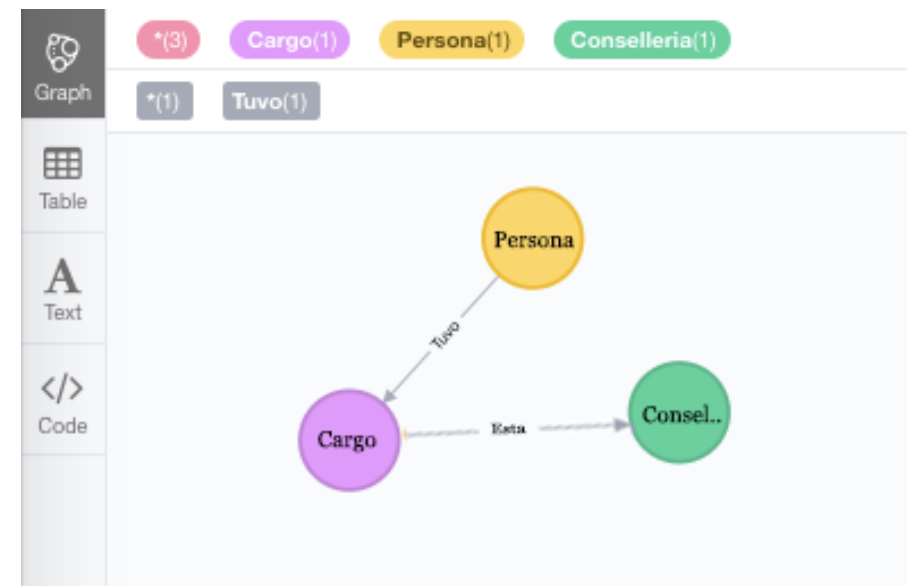
Paso 5: Implementación de los datos

- La estructura del proyecto tendrá:
 - Nodos - registros de datos, son el nombre para los registros de datos (Legislaturas, Consellerías, nombramientos)
 - Relaciones - conectar nodos
 - Propiedades - valores de datos con nombre, atributos que contiene cada nodo.

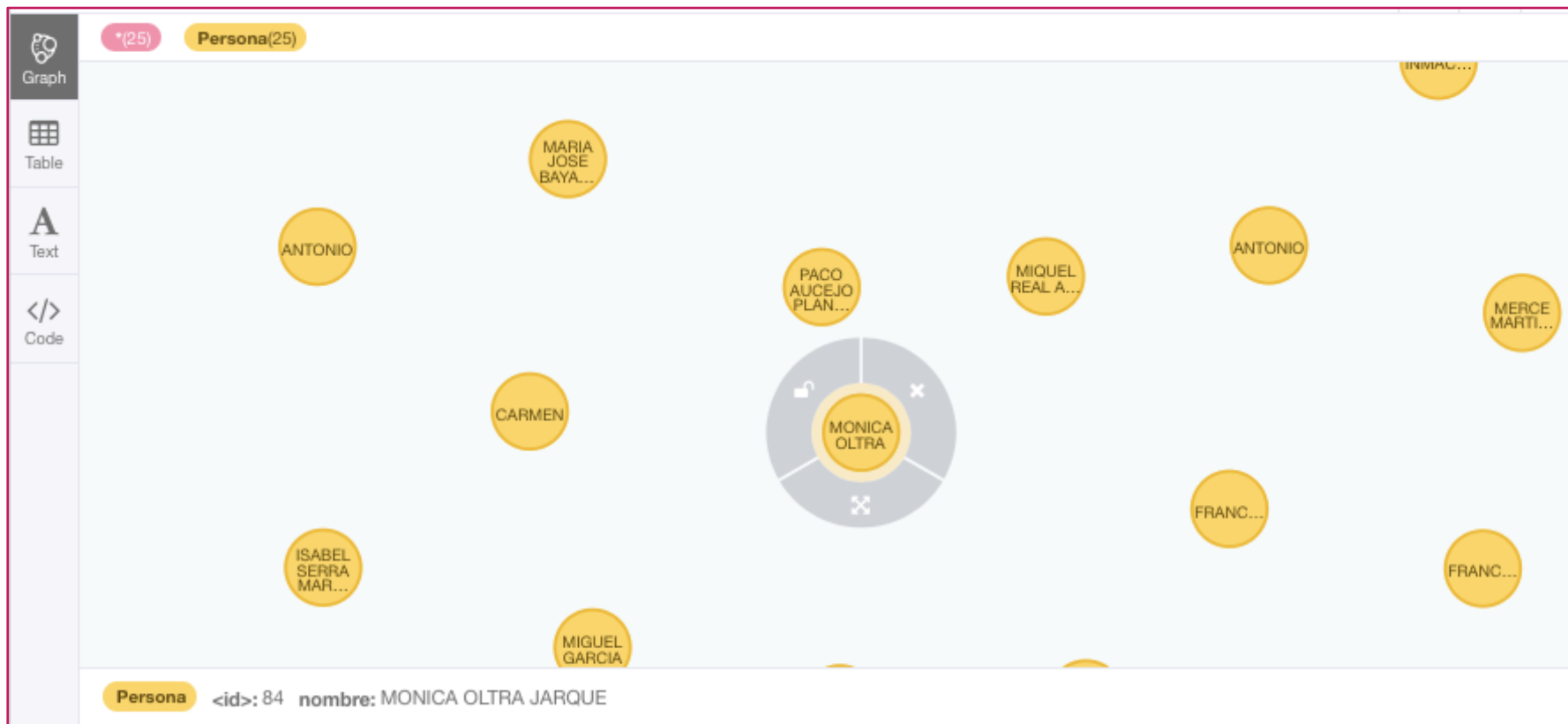
name: 'Ximo Puig'

name: 'Presidente'
conselleria: 'Generalitat Valenciana'

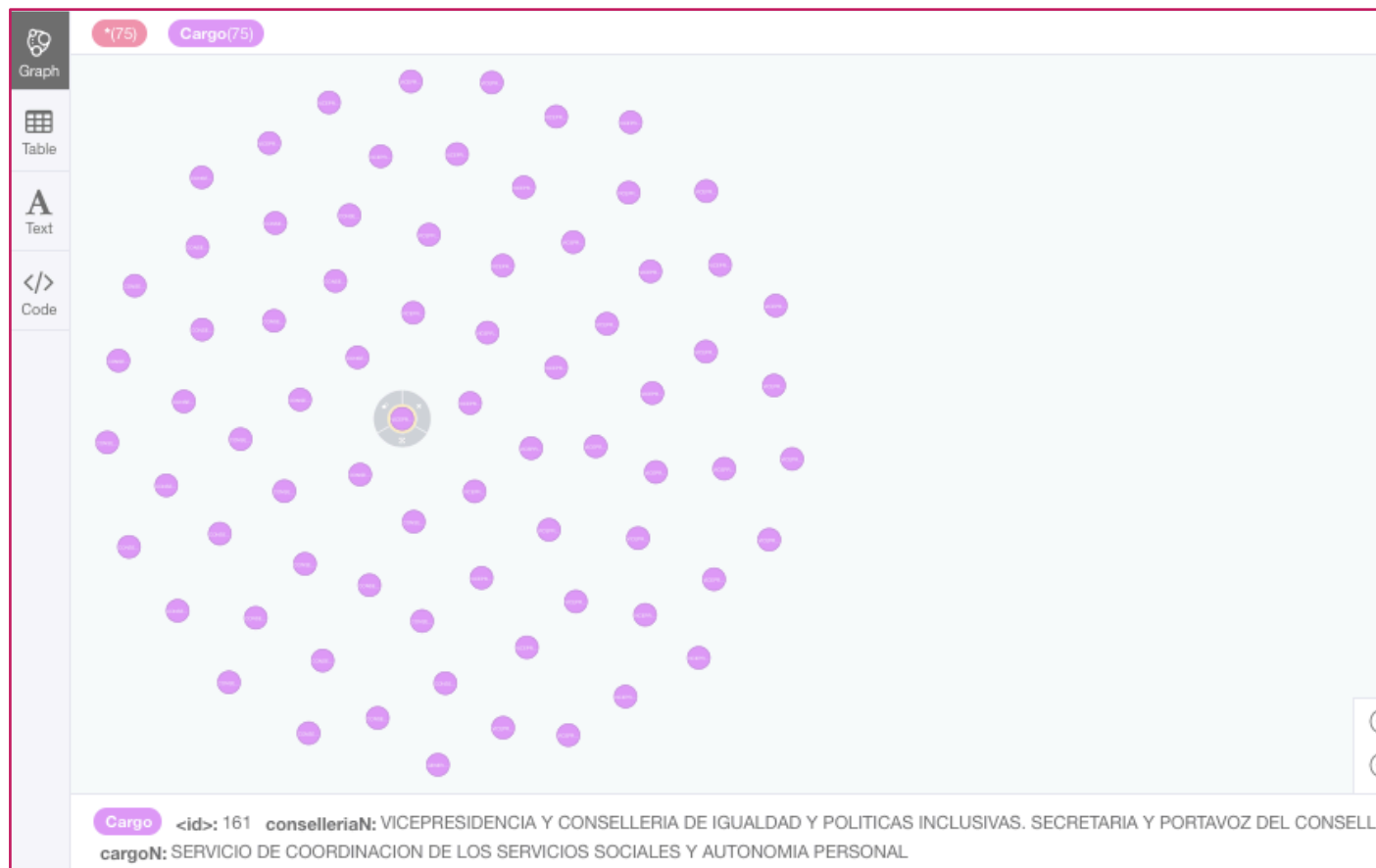
name: 'Conselleria'
ROF: 'Decreto#..'



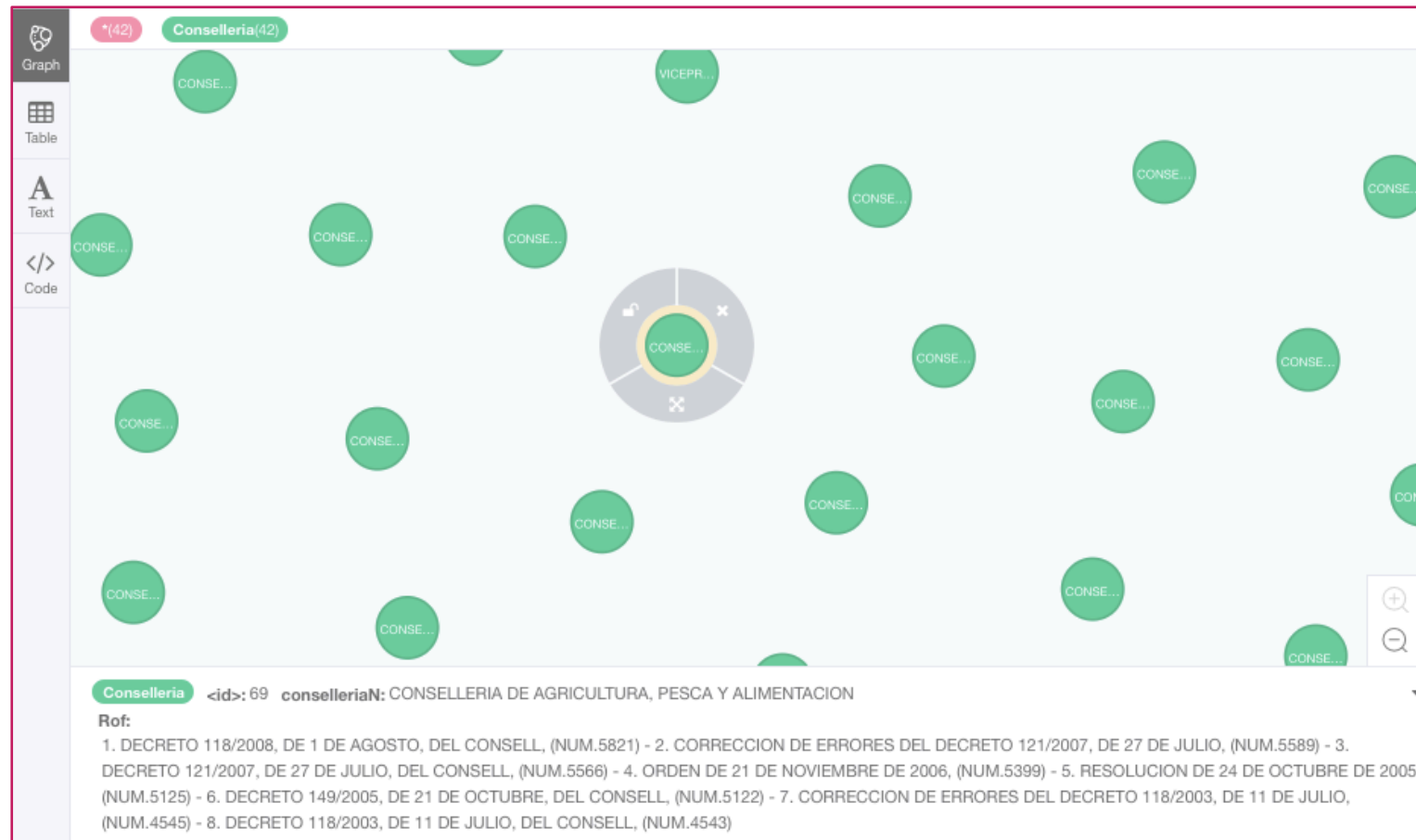
Recolección de datos “Personas”



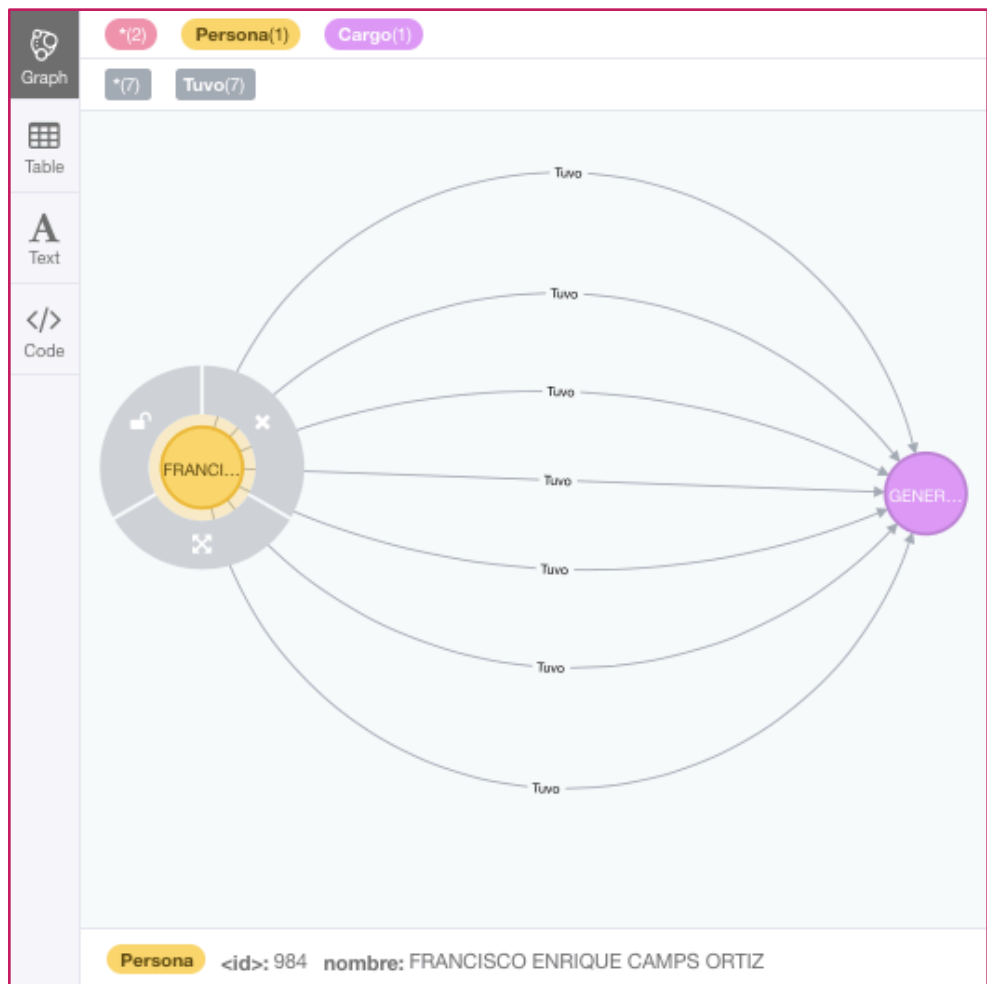
Recolección de datos “Cargos”



Recolección de datos “Consellerías”



Consulta #1



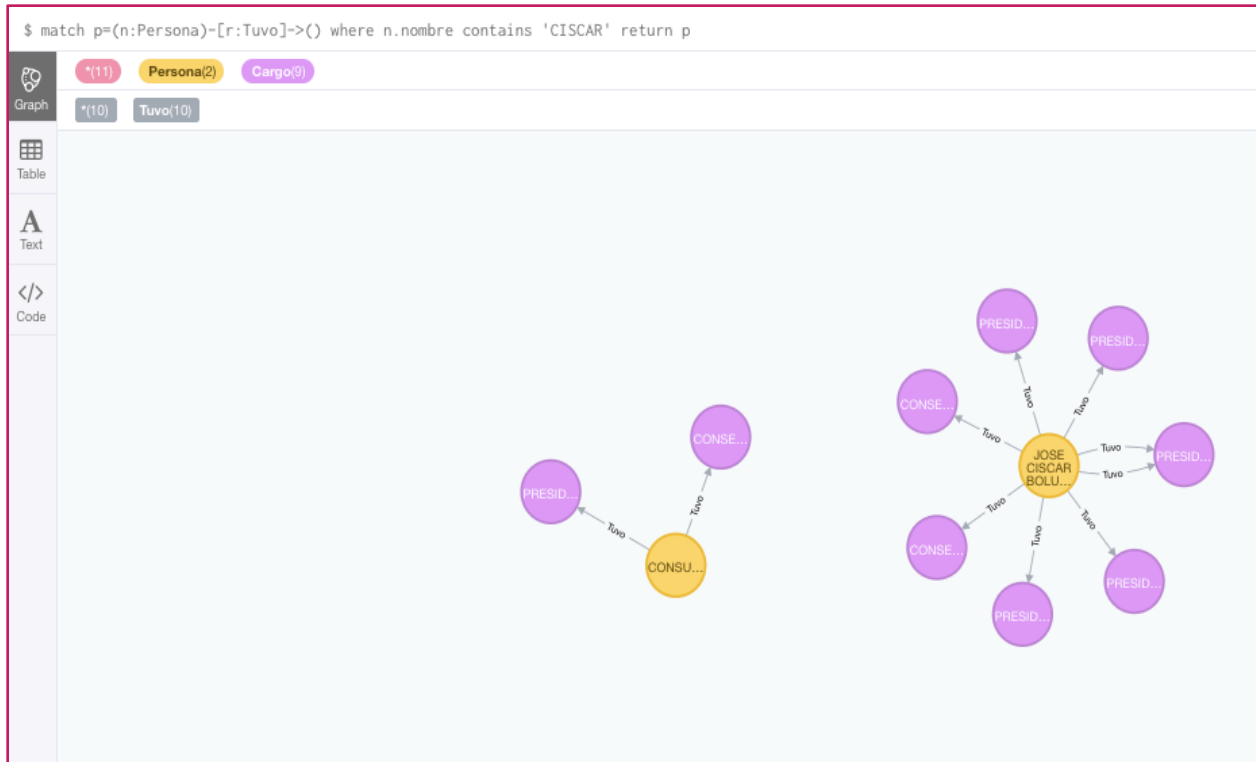
Consulta:

Información de la persona Francisco Enrique Camps Ortiz

Resultado

- Relaciones: Un total de 7.
- Cargos: Presidente de la GVA.
- Fecha de Nombramiento.
- Legislatura y Gobierno que le compete.

Consulta #2



Consulta:
Información de la persona con nombre “CISCAR”

Resultado

- 2 Personas con nombre “CISCAR”
- Relaciones: Un total de 10.
- Cargos: 9

La contratación pública al detalle. Comparador de las contrataciones en las Comunidades Autónomas*

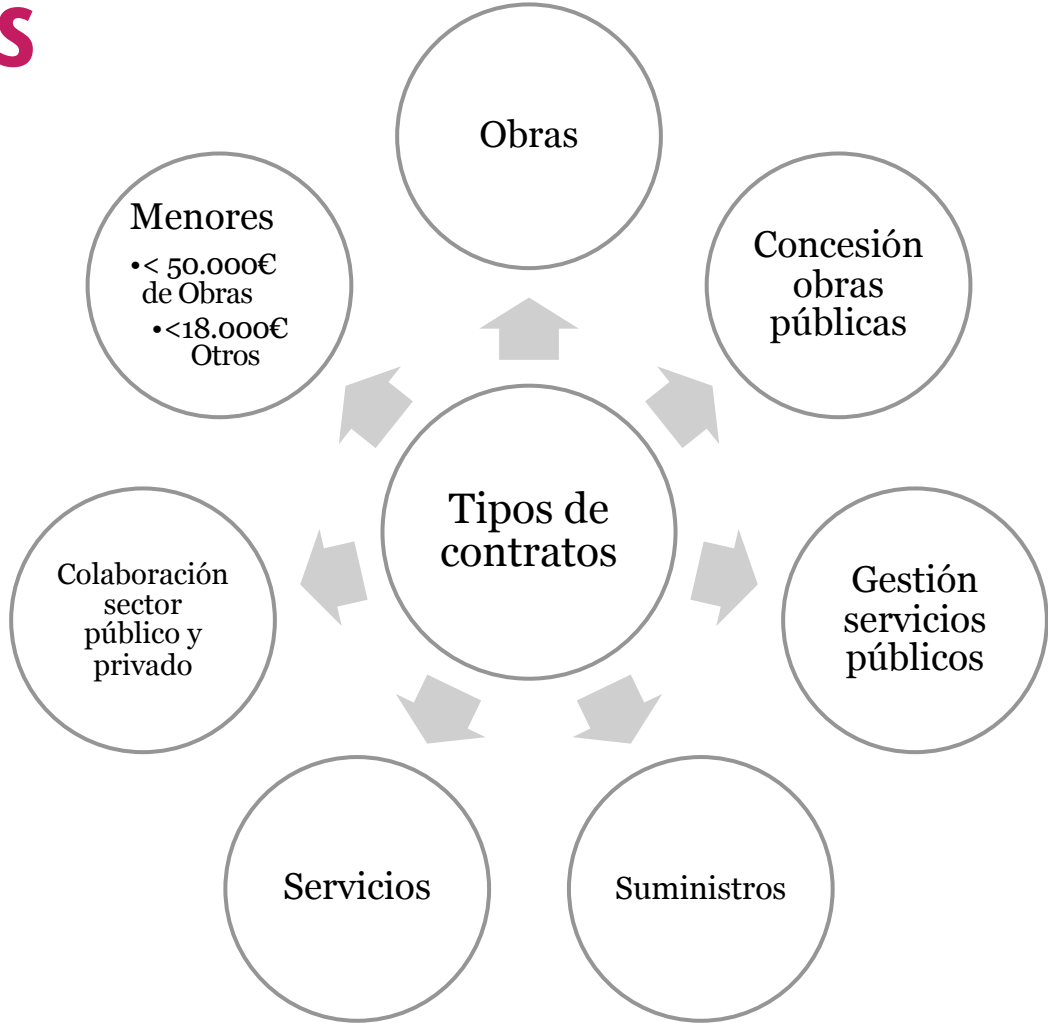
- Sara Diago Gonzalvo

*Comunidades Autónomas equivalente a departamentos en Colombia

Primeros pasos:

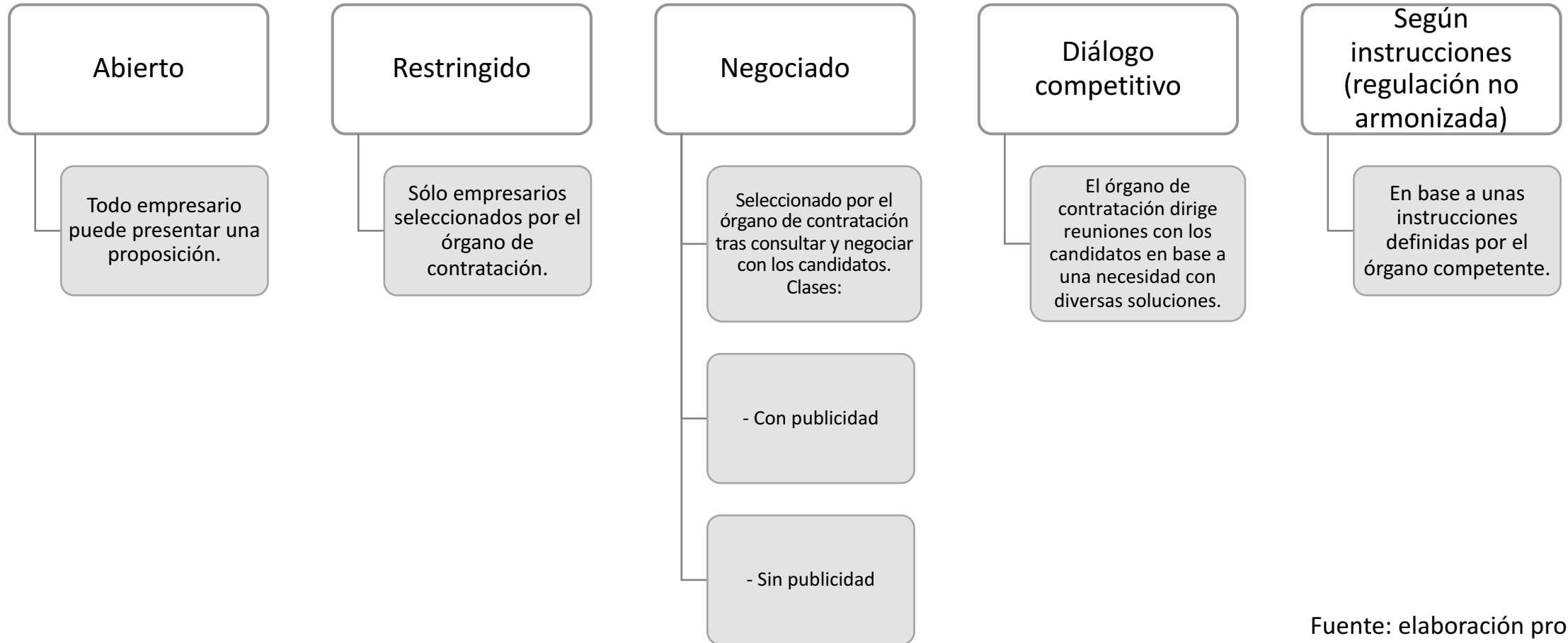
Conocimiento del entorno de las contrataciones

Tipo de contratos



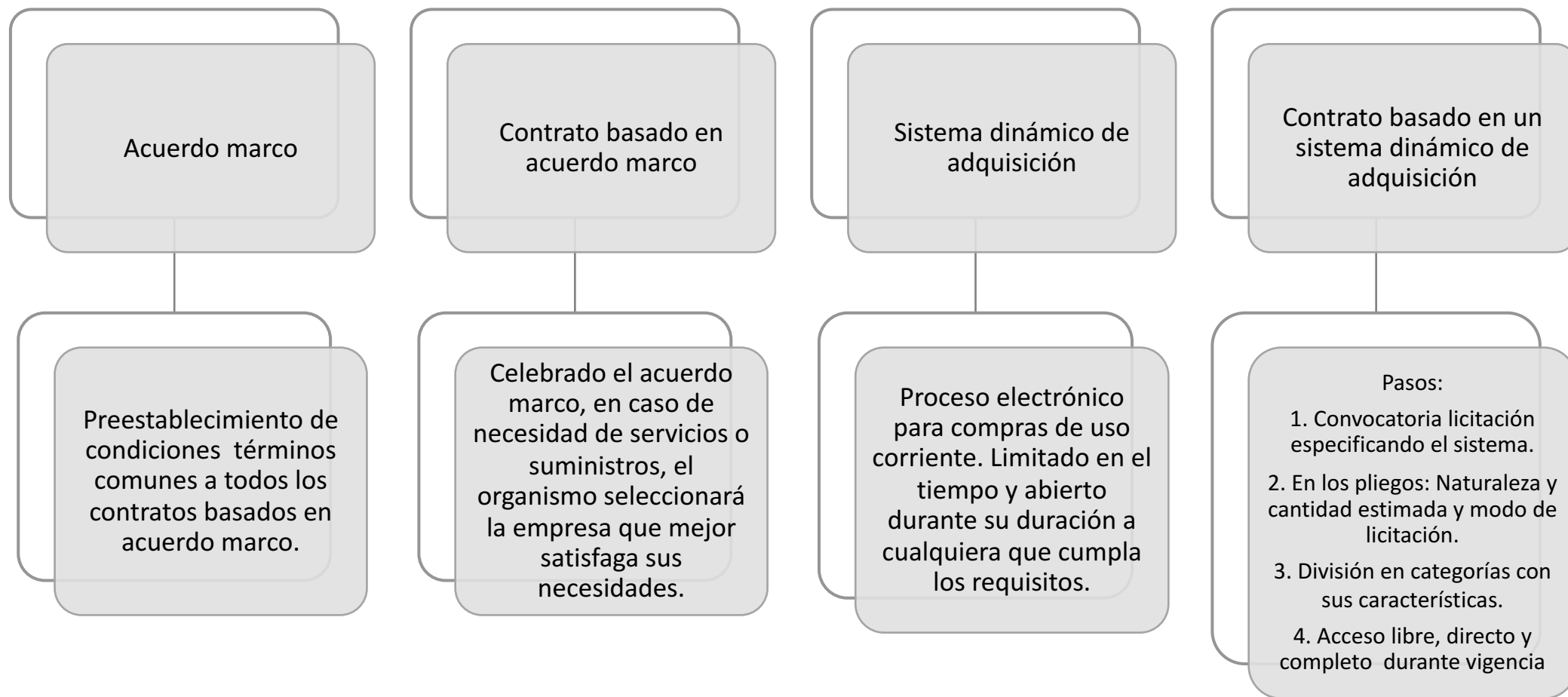
Fuente: elaboración propia

Procedimiento de adjudicación



Fuente: elaboración propia

Sistema de contratación



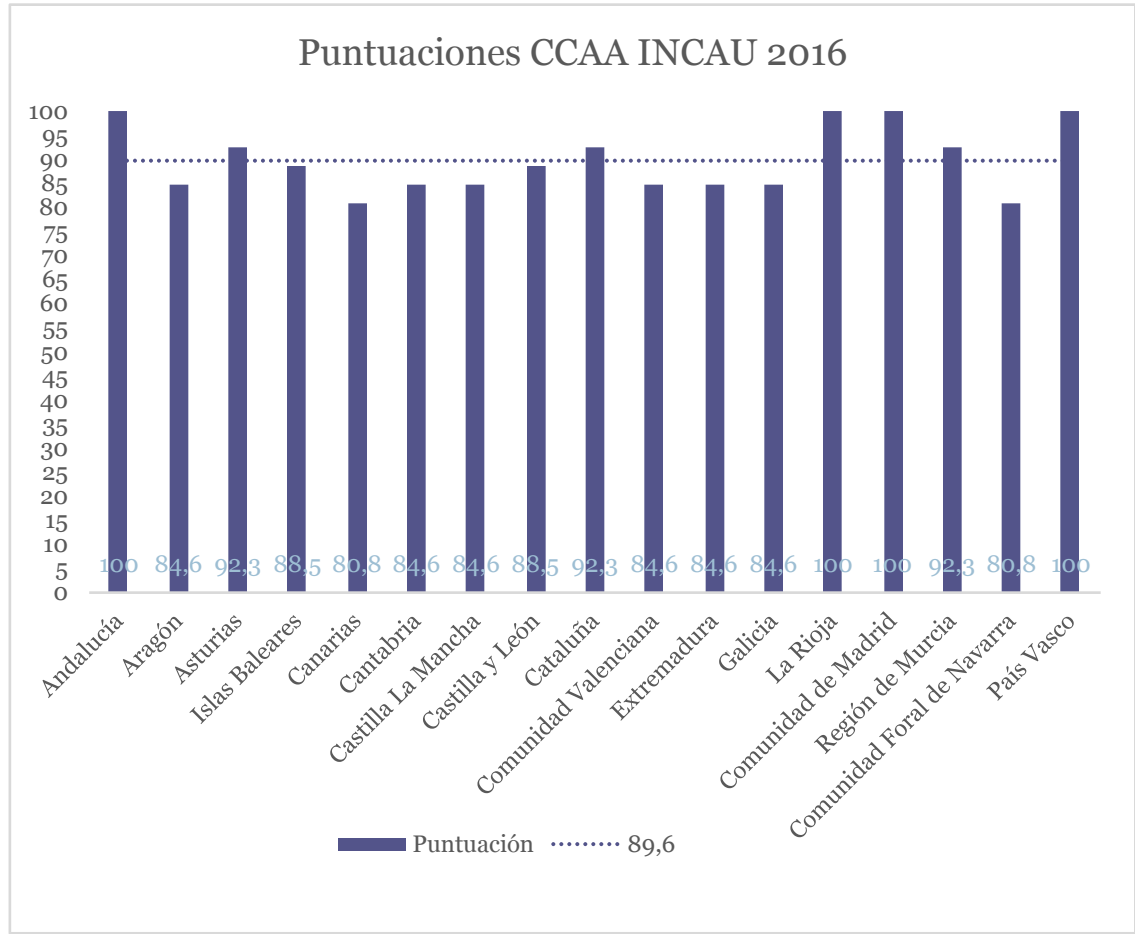
Fuente: elaboración propia

Entorno:

¿Qué entes sociales están implicados en este ámbito?

Transparencia Internacional

Puntuaciones CCAA INCAU 2016



Indicadores	Grado de cumplimiento
1.Contrataciones	
Núm.48 – Publica información contratos	Sí
Núm.49 – Publica información contratos menores	Parcialmente
Núm.50 – Publica modificaciones contratos	Parcialmente
Núm.51 – Información Mesas de Contratación	Parcialmente
Núm.52 – Publica Actas Mesas de Contratación	Parcialmente
Núm.53 – Publica datos estadísticos contratación	Sí
Núm.54 – Informa web publicidad contratos	Sí
1.Convenios y subvenciones	
Núm.55 – Relación convenios suscritos	Sí
Núm.56 – Relación subvenciones y ayudas públicas	Sí
Núm.57 – Publica asignaciones económicas a Grupos Políticos	Sí
Núm.58 – Publica encomiendas de gestión	Sí
1.Suministradores y coste de los servicios	
Núm.59 – Publica proveedores, adjudicatarios y contratistas más importantes	Sí
Núm.60 – Publica coste efectivo de los servicios	Sí

Fuente: elaboración propia en base a los datos obtenidos de TI

Observatorio de la Contratación Pública ObCP

- **Actividad:**

- Impulsar la transparencia, la concurrencia y la integridad en nuestro sistema de contratación pública.
- Mejorar su eficiencia, tanto desde la perspectiva de las administraciones públicas como de los operadores económicos.
- Facilitar el acceso de las PYMES al mercado público.
- Fomentar políticas de I+D+I a través de la contratación pública.
- Promover la participación de la empresa española en el mercado internacional de contratos públicos.
- Favorecer la cooperación y colaboración interadministrativa y público- privada en la contratación pública.



Gobiernos locales:

¿Cuáles son las fuentes de datos que existen en España dentro de este ámbito?

CCAA	Portal de contratación	Portal de transparencia	Portal de Datos Abiertos	Formatos	Clics	Años
Andalucía	100 registros	-	-	RSS	3	2017
Aragón	NO	NO	SÍ	XLS, XML, CSV y JSON	16	2009 - 2017
Asturias	NO	SÍ	SÍ	XLS	5	2015 - 2016
Islas Baleares	NO	NO	Sí	RSS y XML		2017
Canarias	NO	NO	NO	-	-	-
Cantabria	Importe superior a 2 millones de €	SÍ	-	CSV, XLS		2016 - 2017
Castilla La Mancha	Importe superior a 500.000€	-	-	XLS	4	2016
Castilla León	NO	-	SÍ - límite 1 mes	CSV, RSS, XML y JSON	4	2016 - 2017
Cataluña	NO	NO	SÍ	XLS	6	2011 - 2016
Comunidad Valenciana	NO	SÍ	-	CSV	5	2013 - 2017
Extremadura	NO	SÍ	-	XLS	4	2011 - 2017
Galicia	NO	NO	NO	-	-	-
La Rioja	SÍ	SÍ	-	PDF y XLS	3	2016 - 2017
Madrid	SÍ	NO	NO	XLM y XLS		2011 - 2017
Murcia	NO	SÍ	-	XLSX	3	2014 - 2016
Navarra	NO	NO	NO	-	-	-
País Vasco	SÍ	-	-	PDF, XLS, RSS	5	2011 - 2017

Fuente: elaboración propia

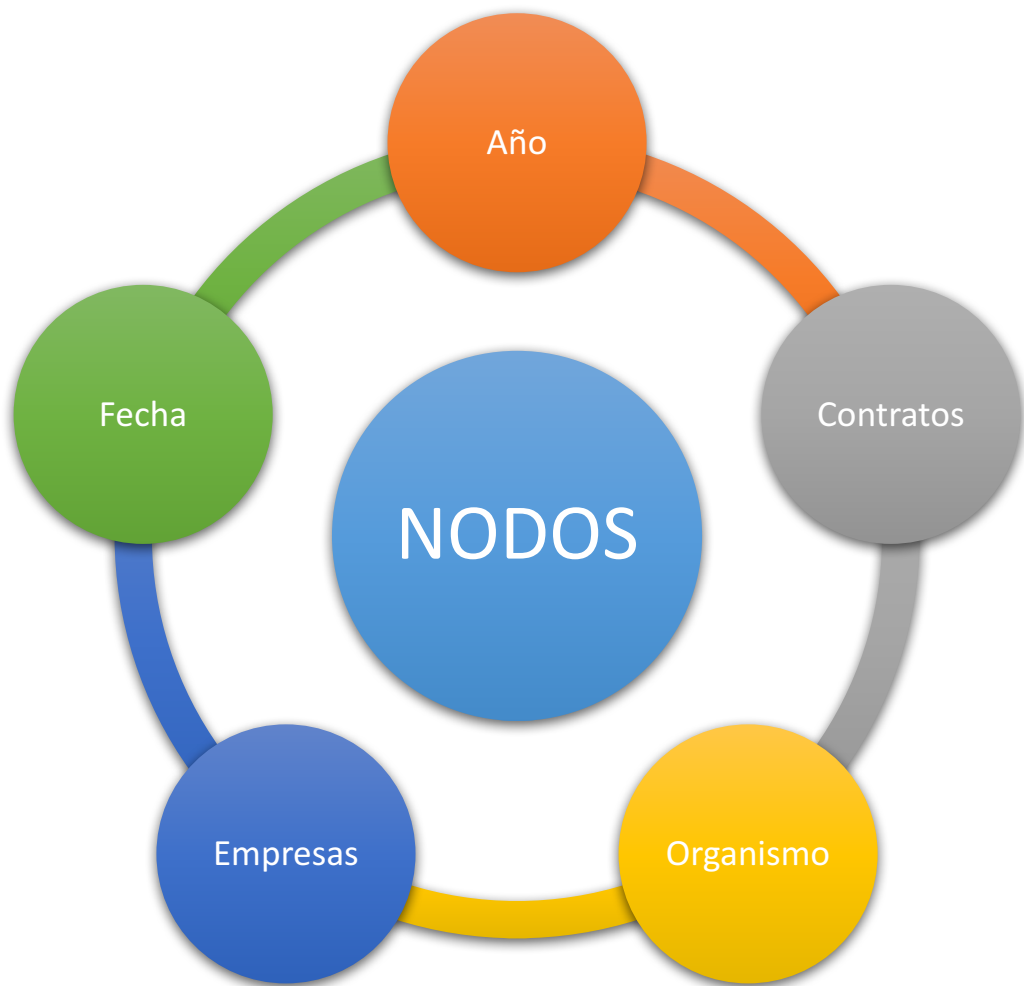
Ejercicio	2017	Núm. Registro:	2601	Núm. Expediente:	CNMY16/INT08/46
Conselleria/Entidad Adjudicadora		Cons. Educación, Investigación, Cultura y		Unidad Administrativa	Ser. d'Infraestructures
Clase contrato		Tipo		Tipo Tramitación	
Menor		No Menor		Servicios	
Forma adjudicación		Sistema de precios		Acuerdo/Contrato marco	
Concurso / Pluralidad de criterios de valoración		Aplicación de honorarios por tarifa		Sí	
Objeto		Xirivella (V) - Servicio Coordinación de Seguridad y Salud en fase de ejecución de las Obras Adecuación/Ampliación del IES Ramón Muntaner con un perfil 16SO+4B+2CCFF(con 4 udes.)+CAF			
Lugar ejecución		VALENCIA		Número de Lotes	
Plazo ejecución		455 Día		Plazo Máximo ejecución	
Importe licitación sin IVA		Importe IVA		Importe Total Licitación	
13.843,82 €		2.907,20 €		16.751,02 €	
Importe adjudicación sin IVA		Importe IVA		Importe Total Adjudicación	
6.750,00 €		1.417,50 €		8.167,50 €	
Importe total final (Extinción por cumplimiento)		Importe final del contrato (Extinción por resolución)			
€		€			
Fecha publicación					
PERFIL	7/10/16	BOE		DOCV	7/10/16
Fecha adjudicación		Fecha formalización		Fecha prevista finalización	
10/4/17		27/4/17		27/7/18	
Variación datos identificación adjudicatario (IDE)		NO		Fecha variación datos identificación	
Variación de plazos y/o importes (MPI)		NO		Fecha variación de plazos y/o importes	
Causa resolución				Fecha de resolución (Extinción por resolución)	
NIF/CIF Adjudicatario		Nombre/Razón Social		Núm Empr invitadas	
B96595897		PREVENCIÓN Y MEDIO AMBIENTE, S.L.-		Núm licitadores	
				22	
Empresas que integran la Unión Temporal de Empresas (UTE)					
NIF/CIF Adjudicatario		Nombre/Razón Social			

Dataset:

- 28.948 registros de contratos de la Generalitat Valenciana.
- 48 columnas con información estructurada.
- Formato CSV.
- Años 2013 a 2017.

Desarrollo tecnológico:

¿Cuál es el esquema de relaciones?



Fuente: elaboración propia

Desarrollo tecnológico:

¿Cuál es la información que podemos extraer de los datos?

Consulta:

¿Cuántos contratos se han licitado por año?

```
$ match (n:Contratos) return n.Tipo as Tipo, count(n.Tipo) as Numero_Contratos order by Numero_Contratos
```

Tipo	Numero_Contratos
"Concesión de obras"	2
"Concesión de Servicios"	3
"Administrativo Especial"	32
"Otros Privados"	33
"Gestión de Servicios Públicos"	47
"Seguros"	65
"Servicios de creación e interpretación artística o literaria"	82
"Servicios de Espectáculos"	370
"Obras"	1939
"Servicios"	10728
"Suministros"	15646

Consulta: Importe por consellería, año y tipo de contrato

Año	Conselleria	Tipo	Importe
"2013"	"Conselleria de Bienestar Social"	"Gestión de Servicios Públicos"	496537778
"2016"	"Conselleria de Sanidad Universal y Salud Pública"	"Suministros"	344836300
"2013"	"Conselleria de Sanidad"	"Gestión de Servicios Públicos"	316244689
"2014"	"Conselleria de Sanidad"	"Suministros"	197578871
"2013"	"Conselleria de Sanidad"	"Servicios"	140605697
"2016"	"Cons. Educación. Investigación. Cultura y Deporte"	"Servicios"	139827726
"2015"	"Conselleria de Sanidad Universal y Salud Pública"	"Suministros"	130467770
"2015"	"Conselleria de Sanidad"	"Suministros"	104030609
"2014"	"Cons. Infraestructuras. Territorio y M. Ambiente"	"Obras"	95572121
"2013"	"Conselleria de Bienestar Social"	"Servicios"	92694865
"2015"	"Conselleria de Bienestar Social"	"Servicios"	83406415
"2015"	"Ciudad de las Artes y de las Ciencias. S.A."	"Otros Privados"	81052224
"2016"	"Vicepresidencia y Conselleria de Igualdad y Políticas Inclusivas"	"Servicios"	78771708
"2014"	"Conselleria de Sanidad"	"Servicios"	72809406
"2013"	"Conselleria de Sanidad"	"Suministros"	70646572
"2014"	"Conselleria de Bienestar Social"	"Servicios"	63524921
"2015"	"Vicepresidencia y Conselleria de Igualdad y Políticas Inclusivas"	"Servicios"	40617448
"2016"	"C. Vivienda. Obras Públicas y Vertebración del Territorio"	"Obras"	36149741
"2015"	"Cons. Infraestructuras. Territorio y M. Ambiente"	"Obras"	32995891
"2016"	"Cons. Educación. Investigación. Cultura y Deporte"	"Obras"	31418207

Consulta: Procedimiento más utilizado por año e importe

Año	Procedimiento	usado	sum(tofloat(re.ImporteTotalLicitación))
"2015"	"Adjudicación directa previa consulta a varias empresas"	4565	56066310.8800003
"2016"	"Adjudicación directa previa consulta a varias empresas"	4465	57209707.450000145
"2014"	"Adjudicación directa previa consulta a varias empresas"	3455	40227500.3800001
"2015"	"Adjudicación directa previa consulta a una empresa"	2419	19785705.109999966
"2016"	"Adjudicación directa previa consulta a una empresa"	2307	18064617.259999983
"2014"	"Adjudicación directa previa consulta a una empresa"	1641	13743063.29000001
"2017"	"Adjudicación directa previa consulta a varias empresas"	1182	14942767.829999989
"2013"	"Adjudicación directa previa consulta a varias empresas"	1146	14204996.989999999
"2014"	"Negociado"	1076	59669311.92
"2016"	"Abierto"	881	476972284.90000015
"2015"	"Abierto"	874	541538669.1199996
"2013"	"Adjudicación directa previa consulta a una empresa"	782	8807967.31
"2014"	"Abierto"	670	489146194.7699999
"2017"	"Adjudicación directa previa consulta a una empresa"	573	4218497.330000002
"2015"	"Negociado"	514	55391198.71999996
"2016"	"Negociado"	495	55105370.43999999
"2013"	"Negociado"	379	39801813.35000001
"2013"	"Abierto"	307	1186659997.7099996
"2017"	"Abierto"	208	65552666.31999999
"2016"	"Basado en acuerdo marco celebrado con varios empresarios"	183	285834715.6400001



La esencia de estos proyectos...

El trabajo en un equipo multidisciplinar que reúne a expertos en distintos ámbitos y la colaboración.



Antonia Ferrer Sapena

Directora / Responsable

Profesor Titular Universidad Politécnica de Valencia.



Fernanda Peset

Investigador Senior / Especialista en Datos de Investigación / Acceso Abierto / Gestión de Datos

Profesor Titular Universidad Politécnica de Valencia.



José Manuel Calabuig

Investigador Senior/ Matemático especialista en Grafos/Bases de datos orientadas a grafos/Visualización de datos

Profesor Titular Universidad Politécnica de Valencia.



Sara Diago Gonzalvo

Community Manager

Alumna del Máster en Gestión de la Información en la Universidad Politécnica de Valencia.



Juan Sebastian Bejarano

WebMaster / Social Media

Alumno del Máster en Gestión de la Información en la Universidad Politécnica de Valencia.



Patricia García Mateo

Aspectos sociales y éticos de los datos / Ontologías

Alumna del Máster en Gestión de la Información en la Universidad Politécnica de Valencia.



German A. Boris B

Analista Big Data / Python

Alumno del Máster en Gestión de la Información en la Universidad Politécnica de Valencia.

¿Cómo y para qué implementar Open Data en las bibliotecas?

- **Las bibliotecas son grandes contenedores de información diversa.**
- **Su gran herramienta: el conocimiento.**
- **La sociedad cambia -> Nuevos formatos**

¿Cuál es la información que necesitamos para impulsar el desarrollo global?

- **No podemos ver la información como única**
- **El conjunto de diversos tipos de información nos proporciona conclusiones acerca de los hechos**
- **Pero... ¡Aquí empieza el debate de los expertos!**

Gracias

Web: <http://www.ctranspa.webs.upv.es>

Twitter: @CtranspaUPV

